Noname manuscript No.
(will be inserted by the editor)

The WASABI Song Corpus and Knowledge Graph for Music Lyrics Analysis

Michael Fell, Elena Cabrio, Maroua Tikat, Franck Michel, Michel Buffa, Fabien Gandon

Received: date / Accepted: date

Abstract We present the WASABI[1] Song Corpus, a large corpus of songs enriched with metadata extracted from music databases on the Web, and resulting from the processing of song lyrics and from audio analysis. More specifically, given that lyrics encode an important part of the semantics of a song, we focus here on the description of the methods we proposed to extract relevant information from the lyrics, such as their structure segmentation, their topics, the explicitness of the lyrics content, the salient passages of a song and the emotions conveyed. The corpus contains 1.73M songs with lyrics (1.41M unique lyrics) annotated at different levels with the output of the above mentioned methods. The corpus labels and the provided methods can be exploited by music search engines and music professionals (e.g. journalists, radio presenters) to better handle large collections of lyrics, allowing an intelligent browsing, categorization and recommendation of songs. We demonstrate the utility and versatility of the WASABI Song Corpus in three concrete application scenarios. Together with the work on the corpus, we present the work achieved to transition the dataset into a knowledge graph, the WASABI RDF Knowledge Graph, and we show how this will enable an even richer set of applications.

Keywords Corpus (Creation, Annotation, etc.), Information Extraction, Information Retrieval, Knowledge Graph, Music and Song Lyrics

Michael Fell
University of Turin.
Corso Svizzera, 185 - 10149 Torino, Italy.
E-mail: mic.fell@gmail.com

Elena Cabrio, Maroua Tikat, Franck Michel, Michel Buffa, Fabien Gandon
Université Côte d'Azur, Inria, CNRS, I3S, France.
930 route des Colles - 06903 Sophia Antipolis CEDEX, France.

[1] **W**eb **A**udio **S**emantic **A**ggregated in the **B**rowser for **I**ndexation

1 Introduction

Let us consider the following scenario: following David Bowie's death, a journalist plans to prepare a radio show about the artist's musical career to acknowledge his qualities. To discuss the topic from different angles, she needs to have at her disposal the artist biographical information to know the history of his career, the song lyrics to know what he was singing about, his musical style, the emotions his songs were conveying, live recordings and interviews, etc. Similarly, streaming professionals such as Deezer, Spotify, Pandora or Apple Music aim at enriching music listening experience with artists' information, to offer suggestions for listening to other songs/albums from the same or similar artists, or automatically determining the emotion felt when listening to a track to propose coherent playlists to the user. To support such scenarios, the need for rich and accurate musical knowledge bases and tools to explore and exploit this knowledge becomes evident.

In the context of the WASABI research project that started in 2017 (funded by the French Research Agency), a two million song database has been built, with metadata on 77k artists, 208k albums, and 2.10M songs [41]. The metadata has been *i)* aggregated, merged and curated from different data sources on the Web, and *ii)* enriched by pre-computed or on-demand analyses of the lyrics and audio data. The WASABI Song Corpus contains songs in 36 different languages, even if the vast majority are in English. As for the songs genres, the most common ones are Rock, Pop, Country and Hip Hop.

Given that lyrics encode an important part of the semantics of a song, in Fell *et al.* [26] we described the methods we have developed to enrich the WASABI Song Corpus with their structure segmentation, the explicitness of the lyrics content, the salient passages of a song, the addressed topics and the emotions conveyed. In the current paper, we extend this previous work in a number of directions by:

– providing additional details on the methods used to extract lyrics metadata.
– presenting two additional application scenarios: scenario B is centered around investigating artists influence (see Section 8.2) and application scenario C focuses on linking the song lyrics to knowledge bases (see Section 8.3).
– describing the formalization, generation and the ongoing publication of the WASABI Song Corpus as an RDF Knowledge Graph (see Section 9).

The paper is organized as follows. Section 2 introduces the WASABI Song Corpus and the metadata initially extracted from music databases on the Web. Section 3 describes the segmentation method we applied to decompose lyrics in their building blocks in the corpus. Section 4 explains the method used to summarize song lyrics, leveraging their structural properties. Section 5 reports on the annotations resulting from the explicit content classifier, while Section 6 describes how information on the emotions are extracted from the lyrics. Section 7 describes the topic modeling we performed on the lyrics.

Section 8 presents three different application scenarios for the corpus. Section 9 describes the formalization, generation and the ongoing publication of the WASABI Song Corpus as an RDF knowledge graph. Section 10 reports on similar existing resources, while Section 11 concludes the paper.

2 The WASABI Song Corpus

In this section, we first describe the steps undertaken to construct the WASABI Song Corpus. We then give key statistics of the obtained dataset along with information on its availability.

2.1 Constructing the WASABI Song Corpus

On the collected song database introduced above, various levels of analysis have been performed, and interactive Web Audio applications have been built on top of the output. For example, the TimeSide analysis and annotation framework have been linked [30] to make on-demand audio analysis possible. In connection with the FAST project[2], an offline chord analysis of 442k songs has been performed, and both an online enhanced audio player [49] and chord search engine [50] have been built around it. A rich set of Web Audio applications and plugins has been proposed [12, 13, 14]. All these metadata, computational analyses and Web Audio applications have now been gathered in one easy-to-use web interface, the WASABI Interactive Navigator[3], illustrated in Figure 1.

We started building the WASABI Song Corpus by collecting for each artist the complete discography, band members with their instruments, time line, equipment they use, and so on. For each song we collected its lyrics from LyricWiki[4], the synchronized lyrics when available[5], the DBpedia abstracts and the categories the song belongs to, e.g. genre, label, writer, release date, awards, producers, artist and band members, the stereo audio track from Deezer, the unmixed audio tracks of the song, its ISRC, bpm and duration.

We matched the song ids from the corpus with the ids from MusicBrainz, iTunes, Discogs, Spotify, Amazon, AllMusic, GoHear, YouTube. Figure 2 illustrates all the data sources we have used to create the dataset. We have also aligned the corpus with the publicly available LastFM dataset[6] which assigns social tags to songs.

As of today, the corpus contains 1.73M songs with lyrics (1.41M unique lyrics). 73k songs have at least an abstract on DBpedia, and 11k have been identified as “classic songs” (they have been number one, or got a Grammy

[2] http://www.semanticaudio.ac.uk
[3] http://wasabi.i3s.unice.fr/
[4] https://en.wikipedia.org/wiki/LyricWiki
[5] from http://usdb.animux.de/
[6] http://millionsongdataset.com/lastfm/

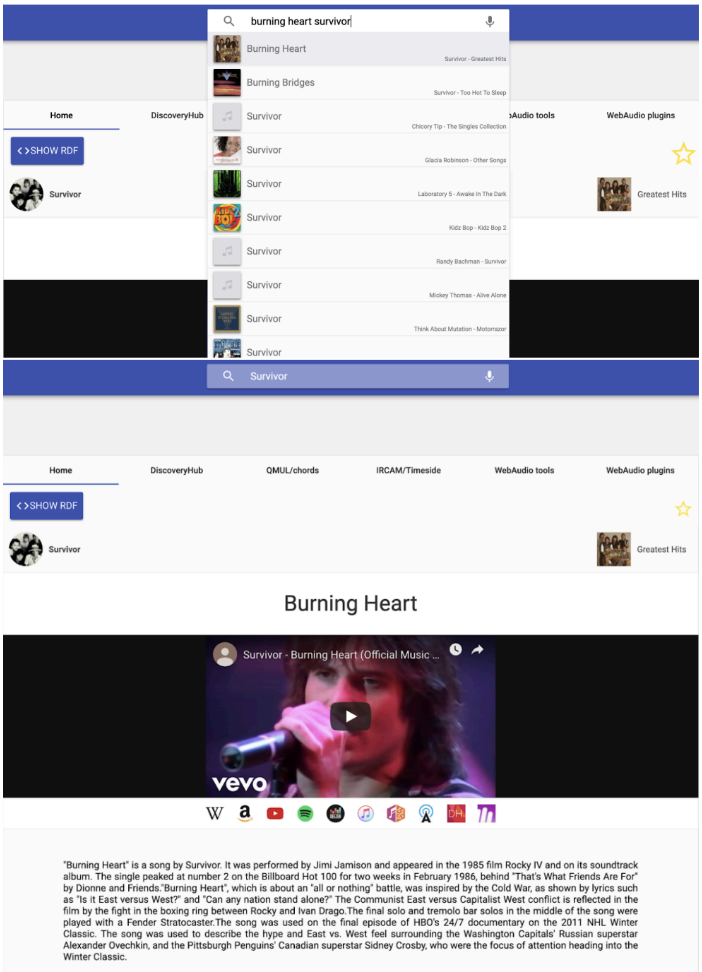


Fig. 1: The WASABI Interactive Navigator [15].

award, or have lots of cover versions). About 2k songs have a multi-track audio version, and on-demand source separation using open-unmix [54] or Spleeter [31] is provided as a TimeSide plugin.

Several Natural Language Processing methods have been applied to the lyrics of the songs included in the WASABI Song Corpus, and various analyses of the extracted information have been carried out. After providing some statistics, the rest of the article describes the different annotations we added to the lyrics of the songs in the dataset. Based on the research we have conducted, the following lyrics annotations are added: lyrical structure (Section 3), summarization (Section 4), explicit lyrics (Section 5), emotion in lyrics (Section 6) and topics in lyrics (Section 7).

Fig. 2: The datasources connected to the WASABI Song Corpus [16].

Table 1 summarizes the most relevant annotations provided in our corpus. Some of those annotation layers are provided for all the 1.73M songs included in the WASABI corpus, while some others apply to subsets of the corpus, due to various constraints described in the next sections.

Annotation	*Labels*	*Description*
Lyrics	1.73M	segments of lines of text
Languages	1.73M	36 different ones
Genre	1.06M	528 different ones
Last FM id	326k	UID
Structure	1.73M	SSM $\in \mathbb{R}^{n \times n}$ (n: length)
Social tags	276k	$\mathbb{S} = \{$rock, joyful, 90s, ...$\}$
Emotion tags	87k	$\mathbb{E} \subset \mathbb{S} = \{$joyful, tragic, ...$\}$
Explicitness	455k	True (85k), False (370k)
Explicitness ♣	715k	True (52k), False (663k)
Summary♣	50k	four lines of song text
Emotion	16k	(valence, arousal) $\in \mathbb{R}^2$
Emotion♣	1.73M	(valence, arousal) $\in \mathbb{R}^2$
Topics♣	1.05M	Prob. distrib. $\in \mathbb{R}^{60}$
Total tracks	2.10M	diverse metadata

Table 1: Most relevant song-wise annotations in the WASABI Song Corpus. Annotations with ♣ are predictions of our models.

2.2 Statistics on the WASABI Song Corpus

This section summarizes key statistics on the corpus, such as the language and genre distributions, the songs coverage in terms of publication years, and then gives the technical details on its accessibility.

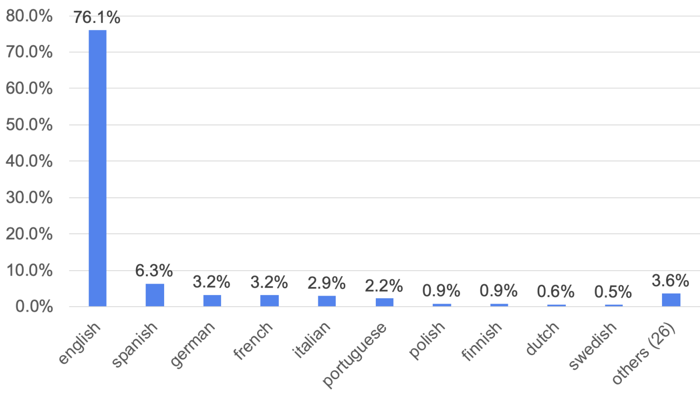
Selection Bias. The amount of lyrics published per decade, per genre, per language etc. is biased by the availability of lyrics for that decade on the crawled web sites. More specifically, the WASABI project initiated the construction of its knowledge base in 2016, at the beginning of the project, taking as its "initial seed" the songs published on the now closed LyricsWikia site, a crowd sourcing wiki manually created by humans. The fact that such source contains more English songs than others, more popular songs, etc. is reflected in the following distributions of our dataset.

Language Distribution. Figure 3a shows the distribution of the ten most frequent languages in our corpus.[7] In total, the corpus contains songs of 36 different languages. The vast majority (76.1%) is English, followed by Spanish (6.3%) and by four languages in the 2-3% range (German, French, Italian, Portuguese). On the bottom end, Swahili and Latin amount to 0.1% (around 2k songs) each.

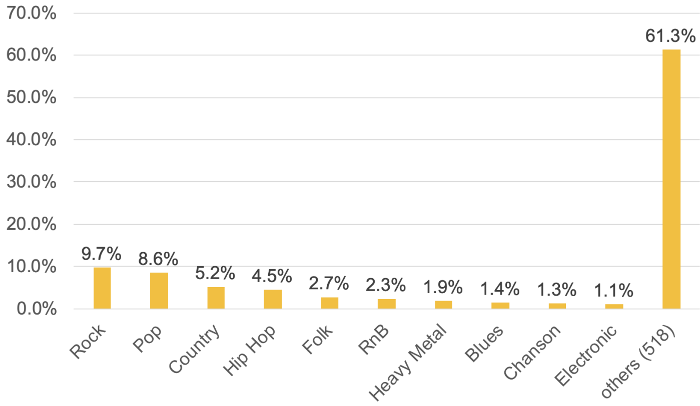
Genre Distribution. In Figure 3b we depict the distribution of the ten most frequent genres in the corpus.[8] In total, 1.06M of the titles are tagged with a genre. It should be noted that the genres are very sparse with a total of 528 different ones. This high number partially stems from the fact that many "subgenres" such as Alternative Rock, Indie Rock, Pop Rock all constitute own genres in the dataset and we take them as is. Specifically, we do not perform any clustering to - for instance - unite *Alternative Rock* with *Rock*. We omit displaying "subgenres" in Figure 3b for clarity. The most common genres are Rock (9.7%), Pop (8.6%), Country (5.2%), Hip Hop (4.5%) and Folk (2.7%).

Publication Year. Figure 3c shows the number of songs published in our corpus, by decade.[9] Over 50% of all songs in the WASABI Song Corpus are from the 2000s or later and only around 10% are from the seventies or earlier. Note one peculiarity, the number of songs in the 2010-2020 decade is lower than in the previous decade. The main reason for this is the initial creation of the dataset in 2016, i.e. our dataset can contain only songs for 6 years of the decade concerned.

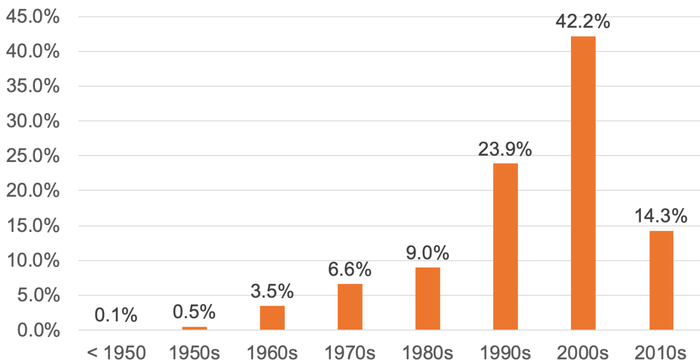
[7] Based on language detection performed on the lyrics.
[8] We take the genre of the album as ground truth since song-wise genres are much rarer.
[9] We take the album publication date as proxy since song-wise labels are too sparse.



(a) Language distribution (100% = 1.73M)



(b) Genre distribution (100% = 1.06M)



(c) Decade of publication distribution (100% = 1.70M)

Fig. 3: Statistics on the WASABI Song Corpus

2.3 Dataset Quality Control

Assessing and ensuring the quality of a large dataset formed by the aggregation of multiple data sources is a difficult and ongoing process. Metadata from various sources can often be inaccurate or contradictory. Multiple hackathons were held between 2017 and 2021, during which a variety of people from different backgrounds and with varying levels of expertise explored the dataset in an open-ended manner. These hackathons have helped identify many recurring errors, conflicts, and problems, and we have written a series of scripts (available in the dataset's Github repository) to fix some of them.

The dataset is intended to be maintained for at least the next three years, as we have new projects and ongoing PhD theses around this work, that will leverage and extend it. More metadata will be added (from the MIR audio analysis of the songs, or obtained by linking the songs to existing midi transcriptions or to their online music scores), and we are developing new metadata quality assessment tools (based on visualizations and inference rules) that will allow the community to better detect and report erroneous, contradictory or missing metadata [42]. On the other hand, the quality of the extracted lyrics metadata has been validated using different methods described in the articles cited in following sections of this paper about lyrics analysis.

2.4 Availability and Accessibility of the WASABI Song Corpus

The WASABI Song Corpus is available and accessible in multiple ways and formats. First of all, it can be searched by end-users through the WASABI Interactive Navigator[10], a rich Web application that relies on a MongoDB back-end server and an Elasticsearch index. It can also be queried by applications in two ways: either using a REST API[11] backed by the MongoDB database, or using our public SPARQL endpoint hosted on a Virtuoso OS triple store[12]. Furthermore, the dataset is also provided as a dump that can be downloaded from Zenodo[13]. The dump contains the dataset in two different formats: the JSON format extracted from our MongoDB back-end, and the RDF knowledge graph that is described in details in section 9. In all cases (web application, API/SPARQL programmatic access, dumps) all the metadata is publicly available under a Creative Commons Attribution-Non Commercial license.

2.5 Getting Access to Copyrighted Content

It is important to note that during the project we had access to copyrighted content - song lyrics and audio files - that we are not allowed to include in

[10] `http://wasabi.i3s.unice.fr/`
[11] `https://wasabi.i3s.unice.fr/apidoc/`
[12] `http://wasabi.inria.fr/sparql`
[13] `https://doi.org/10.5281/zenodo.5603369`

the publicly released dataset. Nevertheless, significant work has been done on the analysis of the lyrics and the results are available on the project's GitHub (metadata, ML models, Python scripts and Jupyter notebooks). These results open the way to many scientific uses by the NLP community.

Notice that an online data source we used for song lyrics at the beginning of the WASABI project (2017-2021), LyricsWikia, is now offline[14]. However, it is still possible for researchers to obtain the full lyrics we used using commercial APIs such as MusixMatch[15]. Moreover, if you contact the authors, it is possible to obtain our unrestricted lyrics files under conditions of confidentiality and scientific use only. As for the audio content of the songs, 30-second clips are accessible via the public Deezer Audio API or full-length content through the YouTube API.

3 Lyrics Structure Annotations

Generally speaking, lyrics structure segmentation consists of two stages: text segmentation to divide lyrics into segments, and semantic labeling to label each segment with a structure type (e.g. Intro, Verse, Chorus).

In [27] we proposed a method to segment lyrics based on their repetitive structure in the form of a self-similarity matrix (SSM). Figure 4 shows a line-based SSM for the song text written on top of it[16]. The lyrics consist of seven segments and shows the typical repetitive structure of a Pop song. The main diagonal is trivial, since each line is maximally similar to itself. Notice further the additional diagonal stripes in segments 2, 4 and 7; this indicates a repeated part, typically the chorus. Based on the simple idea that eyeballing an SSM will reveal (parts of) a song's structure, we proposed a Convolutional Neural Network architecture that successfully learned to predict segment borders in the lyrics when "looking at" their SSM. Table 2 shows the genre-wise results we obtained using our proposed architecture. One important insight was that more repetitive lyrics as often found in genres such as Country and Punk Rock are much easier to segment than lyrics in Rap or Hip Hop which often do not even contain a chorus. We found that in many cases, where the lyrics do not reveal the segment structure, the audio of the song can come to aid and inform the segmentation model [29]. Leveraging a corpus of synchronized text-audio representations, we showed that taking into account both modalities - the song audio and the song lyrics - improved the segmentation performance.[17]

In the WASABI Interactive Navigator, the line-based SSM of a song text can be visualized. It is toggled by clicking on the violet-blue square on top of

[14] `https://en.wikipedia.org/wiki/LyricWiki`

[15] `https://developer.musixmatch.com/`, here is a sample code for lyrics retrieval using this API at `https://jsbin.com/joyifojuva/edit`

[16] `https://wasabi.i3s.unice.fr/#/search/artist/Britney%20Spears/album/In%20The%20Zone/song/Everytime`

[17] Obtained f-scores ranged between 70.8% for text-based and 75.3% for text-audio-based models.

Fig. 4: Structure of the lyrics of "Everytime" by Britney Spears as displayed in the WASABI Interactive Navigator.

Genre	P	R	F_1
Rock	73.8	57.7	64.8
Hip Hop	71.7	43.6	<u>54.2</u>
Pop	73.1	61.5	66.6
RnB	71.8	60.3	65.6
Alternative Rock	76.8	60.9	67.9
Country	74.5	66.4	**70.2**
Hard Rock	76.2	61.4	67.7
Pop Rock	73.3	59.6	65.8
Indie Rock	80.6	55.5	65.6
Heavy Metal	79.1	52.1	63.0
Southern Hip Hop	73.6	34.8	<u>47.0</u>
Punk Rock	80.7	63.2	**70.9**
Alternative Metal	77.3	61.3	68.5
Pop Punk	77.3	68.7	**72.7**
Gangsta Rap	73.6	35.2	<u>47.7</u>
Soul	70.9	57.0	63.0

Table 2: Lyrics segmentation performances across musical genres in terms of Precision (P), Recall (R) and F_1 in %. Performances on genres with less repetitive and highly repetitive structures are underlined and in bold, respectively.

the song text. For a subset of songs the color opacity indicates how repetitive and representative a segment is, based on the fitness metric that we proposed in [25]. Note how in Figure 4, the segments 2, 4 and 7 are shaded more darkly than the surrounding ones. As highly fit (opaque) segments often coincide with a chorus, this is a first approximation of chorus detection. Given the variability in the set of structure types provided in the literature according to different genres [55, 10], rare attempts have been made in the literature to achieve a more complete semantic labeling, labeling the lyrics segments as Intro, Verse, Bridge, Chorus etc.

For each song text, we provide an SSM based on a normalized character-based edit distance[18] on two levels of granularity, to enable other researchers to work with these structural representations: line-wise similarity and segment-wise similarity.

4 Lyrics Summary

Given the repeating forms, peculiar structure and other unique characteristics of song lyrics, in [25] we introduced a method for extractive summarization of lyrics that takes advantage of these additional elements to more accurately identify relevant information in song lyrics. More specifically, it relies on the intimate relationship between the audio and the lyrics. The so-called audio thumbnails, snippets of usually 30 seconds of music, are a popular means to summarize a track in the audio community. The intuition is this: the more repeated and the longer a part, the better it represents the song. We transferred an audio thumbnailing approach to our domain of lyrics and showed that adding the thumbnail improves summary quality. We evaluated our method on 50k lyrics belonging to the top 10 genres of the WASABI Song Corpus and according to qualitative criteria such as *Informativeness* and *Coherence*. Figure 5 shows our results for different summarization models. Our model $\mathbb{R}ank\mathbb{T}opic\mathbb{F}it$, which combines graph-based, topic-based and thumbnail-based summarization, outperforms all other summarizers. We further find that the genres RnB and Country are highly overrepresented in the lyrics sample with respect to the full corpus, indicating that songs belonging to these genres are more likely to contain a chorus. Finally, Figure 6 shows an example summary of four lines length obtained with our proposed $\mathbb{R}ank\mathbb{T}opic\mathbb{F}it$ method. It is toggled in the WASABI Interactive Navigator by clicking on the green square on top of the song text.

The four-line summaries of 50k English song lyrics used in our experiments are freely available within the WASABI Song Corpus; the Python code of the applied summarization methods is also available[19].

[18] In our segmentation experiments we found this simple metric to outperform more complex metrics that take into account the phonetics or the syntax.

[19] `https://github.com/TuringTrain/lyrics_thumbnailing`

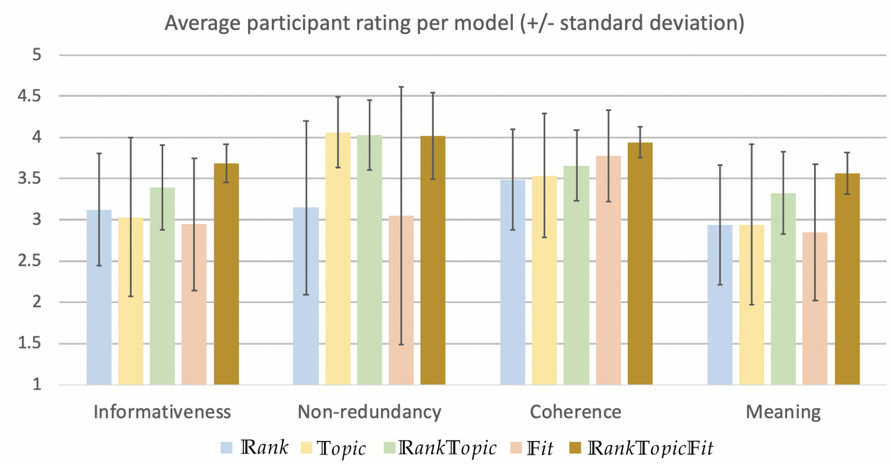


Fig. 5: Human ratings per summarization model (five point Likert scale). Models are $\mathbb{R}ank$: graph-based, $\mathbb{T}opic$: topic-based, $\mathbb{F}it$: thumbnail-based, and model combinations.

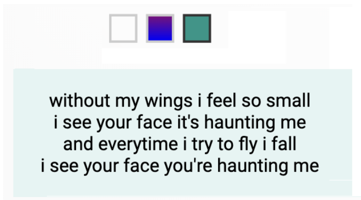


Fig. 6: Summary of the lyrics of "Everytime" by Britney Spears as displayed in the WASABI Interactive Navigator.

5 Explicit Language in Lyrics

On audio recordings, the Parental Advisory Label is placed in recognition of profanity and to warn parents of material potentially unsuitable for children. Nowadays, such labeling is carried out mainly manually on a voluntary basis, with the drawbacks of being time consuming and therefore costly, error prone and partly a subjective task. In [24] we have tackled the task of automated explicit lyrics detection, based on the songs carrying such a label. We compared automated methods ranging from dictionary-based lookup to state-of-the-art deep neural networks to automatically detect explicit contents in English lyrics. More specifically, the dictionary-based methods rely on a swear word dictionary D_n which is automatically created from example explicit and clean lyrics. Then, we use D_n to predict the class of an unseen song text in one of two ways: (i) the *Dictionary Lookup* simply checks if a song text contains words from D_n. (ii) the *Dictionary Regression* uses a bag of words (BOW) made from D_n as the feature set of a logistic regression classifier. In

Model	P	R	F_1
Majority Class	45.0	50.0	47.4
Dictionary Lookup	78.3	76.4	77.3
Dictionary Regression	76.2	81.5	78.5
Tf-idf BOW Regression	75.6	81.2	78.0
TDS Deconvolution	81.2	78.2	79.6
BERT Language Model	84.4	73.7	77.7

Table 3: Performance comparison of our different models. Precision (P), Recall (R) and f-score (F_1) in %.

the *Tf-idf BOW Regression* the BOW is expanded to the whole vocabulary of a training sample instead of only the explicit terms. Furthermore, the model *TDS Deconvolution* is a deconvolutional neural network [56] that estimates the importance of each word of the input for the classifier decision. In our experiments, we worked with 179k lyrics that carry gold labels provided by Deezer (17k tagged as explicit) and obtained the results shown in Figure 3. We found the very simple *Dictionary Lookup* method to perform on par with much more complex models such as the *BERT Language Model* [21] as a text classifier. Our analysis revealed that some genres are highly overrepresented among the explicit lyrics. Inspecting the automatically induced explicit words dictionary reflects that genre bias. The dictionary of 32 terms used for the dictionary lookup method consists of around 50% of terms specific to the Rap genre, such as glock, gat, clip (gun-related), thug, beef, gangsta, pimp, blunt (crime and drugs). Finally, the terms holla, homie, and rapper are obviously not swear words, but highly correlated with explicit content lyrics.

Our corpus contains 52k tracks labeled as explicit and 663k clean (not explicit) tracks[20]. We have trained a classifier (77.3% f-score on test set) on the 438k English lyrics which are labeled and classified the remaining 455k previously untagged English tracks. We provide both the predicted labels in the WASABI Song Corpus and the trained classifier to apply it to unseen text.

6 Emotional Description

In sentiment analysis the task is to predict if a text has a positive or a negative emotional valence. In the recent years, a transition from detecting sentiment (positive vs. negative valence) to more complex formulations of emotion detection (e.g. joy, fear, surprise) [45] has become more visible; even tackling the problem of emotion in context [18]. One family of emotion detection approaches is based on the valence-arousal model of emotion [52], locating every emotion in a two-dimensional plane based on its valence (positive vs. negative) and arousal (aroused vs. calm).[21] Figure 7 is an illustration of the valence-arousal model of Russell and shows exemplary where several emotions such as

[20] Labels provided by Deezer. Furthermore, 625k songs have a different status such as unknown or censored version.

[21] Sometimes, a third dimension of dominance is part of the model.

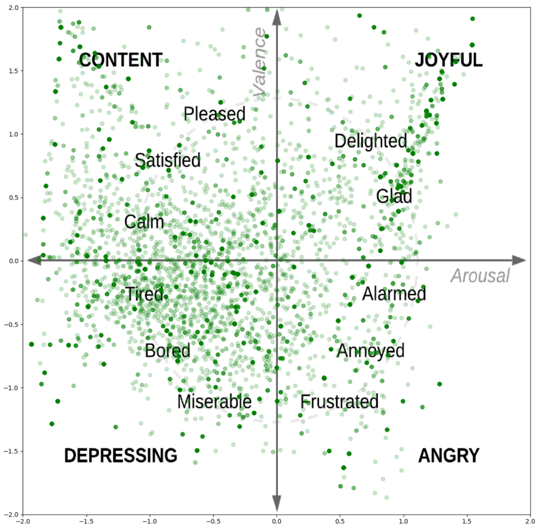


Fig. 7: Emotion distribution in the corpus in the valence-arousal plane.

joyful, angry or calm are located in the plane. Manually labeling texts with multi-dimensional emotion descriptions is an inherently hard task. Therefore, researchers have resorted to distant supervision, obtaining gold labels from social tags from LastFM. These approaches [33, 17] define a list of social tags that are related to emotion, then project them into the valence-arousal space using an emotion lexicon [57, 44].

Recently, Deezer made valence-arousal annotations for 18,000 English tracks available[22] [20], and we transferred the valence-arousal annotations of Deezer to our songs. In Figure 7 the green dots visualize the emotion distribution of these songs.[23] Based on their annotations, we train an emotion regression model using BERT, with an evaluated 0.44/0.43 Pearson correlation/Spearman correlation for valence and 0.33/0.31 for arousal on the test set.

We integrated Deezer's labels into our corpus and also provide the valence-arousal predictions for the 1.73M tracks with lyrics [23]. We also provide the LastFM social tags (276k) and emotion tags (87k entries) to help researchers with building variants of emotion recognition models.

7 Topic Modelling

We built a topic model on the lyrics of our corpus using Latent Dirichlet Allocation (LDA) [9]. We determined the hyperparameters α, η and the topic

[22] `https://github.com/deezer/deezer_mood_detection_dataset`

[23] Depiction without scatter plot taken from [47]

Fig. 8: Topic War

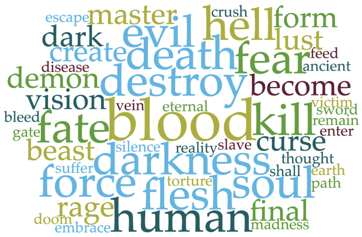


Fig. 9: Topic Death

Fig. 10: Topic Love

Fig. 11: Topic Family

Fig. 12: Topic Money

Fig. 13: Topic Religion

count such that the coherence was maximized on a subset of 200k lyrics. We then trained a topic model of 60 topics on the unique English lyrics (1.05M).

We manually labeled a number of more recognizable topics to characterize their semantic content (as War, Death, Love). Figures 9-13 illustrate these topics with word clouds[24] of the most characteristic words per topic. For instance, the topic Money contains words of both the field of earning money (job, work, boss, sweat) as well as spending it (pay, buy). The topic Family is both about the people of the family (mother, daughter, wife) and the land (sea, valley, tree) [23]. We provide the topic distribution of our LDA topic model for each song and make available the trained topic model to enable its application to unseen lyrics.

[24] made with `https://www.wortwolken.com/`

8 Exploiting and Reasoning over the Lyrics Annotation Layers: Application Scenarios

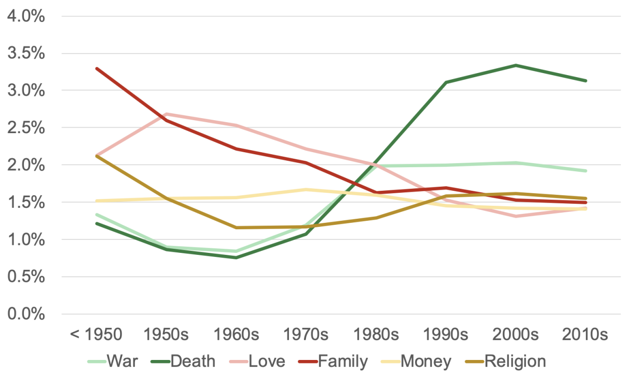
Following a user-oriented approach, in the earlier stages of the WASABI project, a number of potential users of the WASABI Song Corpus were interviewed (as musicologists, journalists, artists), so as to design a set of motivating use-case scenarios. In this section, we analyze them from an NLP perspective, and we show how exploiting and reasoning over the different annotation layers resulting from the application to the song lyrics of the NLP methods we presented could open a wide range of possibilities of analysis on the artists and their musical productions, from different perspectives.

8.1 Application Scenario A: Diachronic Corpus Analysis

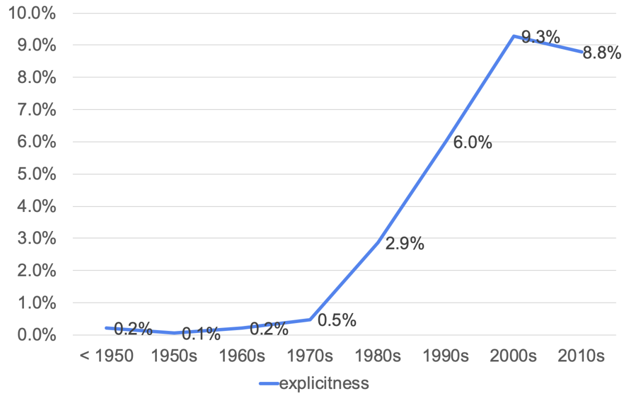
An important task for musicologists is to analyze the discography of a particular artist. Nearly all artists have different sounds - and touch at different topics in their lyrics - over time, but some notably have more drastic changes than others. Some of these changes might just be because of more experience and better production, but some extend beyond that. Digging into them can provide interesting hints on the artists' lives, and on the historical context inspiring them. Adding the time dimension to the metadata we extracted from the lyrics, we can indeed visualize how they evolved during the life of an artist and ultimately try *i)* to understand what caused certain ruptures (for example in the subjects covered, or in the verse/chorus cut of the songs), and *ii)* to study variations in the compositions as the improved/decreased complexity of the songs (audio, texts).

In the rest of this section, we focus on the change over time, providing a diachronic corpus analysis as follows: we examine the changes in the annotations over the course of time, by grouping the corpus into decades of songs according to the distribution shown in Figure 3c. After providing this overview analysis for the whole corpus and therefore for various artists, we close the section by focusing on the development of a specific artist, namely David Bowie.

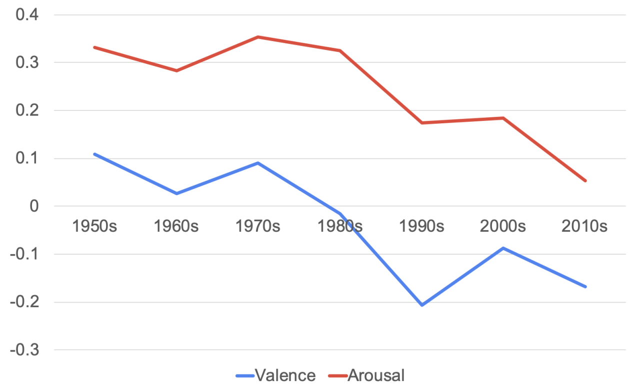
Changes in Topics. The importance of certain topics has changed over the decades, as depicted in Figure 14a. Some topics have become more important, others have declined, or stayed relatively the same. We define the importance of a topic for a decade of songs as follows: first, the LDA topic model trained on the full corpus gives the probability of the topic for each song separately. We then average these song-wise probabilities over all songs of the decade. For each of the cases of growing, diminishing and constant importance, we display two topics. The topics War and Death have appreciated in importance over time. This is partially caused by the rise of Heavy Metal in the beginning of the 1970s, as the vocabulary of the Death topic is very typical for the genre (see for instance the "Metal top 100 words" in [28]). We measure a decline in the importance of the topics Love and Family. The topics Money and Religion seem to be evergreens as their importance stayed rather constant over time.



(a) Evolution of topic importance



(b) Evolution of explicit content lyrics



(c) Evolution of emotion

Fig. 14: Evolution of different annotations during the decades

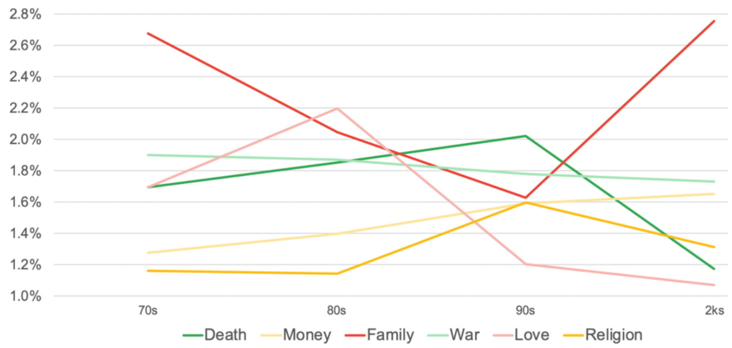


Fig. 15: Topic importance in David Bowie songs over time

Changes in Explicitness. We find that newer songs are more likely being tagged as having explicit content lyrics. Figure 14b shows our estimates of explicitness per decade, the ratio of songs in the decade tagged as explicit to all songs of the decade. Note that the Parental Advisory Label was first distributed in 1985 and many older songs may not have been labeled retroactively. The depicted evolution of explicitness may therefore overestimate the "true explicitness" of newer music and underestimate it for music before 1985.

Changes in Emotion. We estimate the emotion of songs in a decade as the average valence and arousal of songs of that decade. We find songs to decrease both in valence and arousal over time. This decrease in positivity (valence) is in line with the diminishment of positively connoted topics such as Love and Family and the appreciation of topics with a more negative connotation such as War and Death.

As an interesting anecdotal evidence, we find the singer David Bowie to write more about the topic Family in his very young age (70s), to then put more weight on a topic such as Love (80s), to later return to an emphasis in Family in his old days (2000s). We illustrate the topic importance trajectories in Figure 15.

While in this section we have shown the application of the diachronic analysis on the corpus as a whole, and on one specific artist, it can actually be used also to find similarities among artists with regard to the evolution of the topics and emotions addressed by their lyrics, or sharing the same kind of song structures.

8.2 Application Scenario B: Artists Influence

Discovering relations between artists to track how they co-produce, co-write, and ultimately influence each other, is another interesting task to research, and among the needs expressed by journalists and musicologists. We operationalize

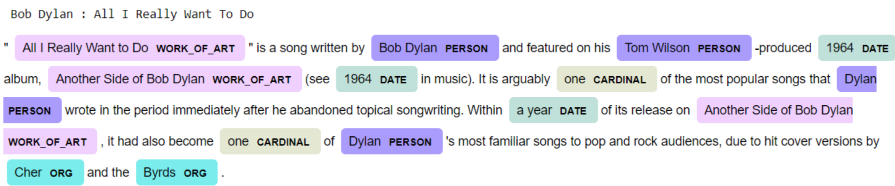


Fig. 16: Example abstract showing the artist *Cher* being mentioned in a *Bob Dylan* song.

this problem in the following, by creating a social network of the artists in our corpus, to unveil connections between them in the WASABI Song Corpus. We base our analysis on the available song abstracts extracted from DBpedia. These are short texts describing the circumstances in which a song was written, composed, collaborations that lead to it, or more generally, relevant context information associated with the song. Central to our interest, we find that inside such song abstracts, other artists related to the song in some way are mentioned. Figure 16 shows an example abstract of a song by the artist *Bob Dylan* in which another artist, *Cher*, is mentioned. In this example, Cher is mentioned since she covered the song.

To build such a social network of artists, we perform Named Entity Recognition on the song abstracts using the spaCy [32] library.[25] As Named Entity Recognition is language-specific, in this work we limit ourselves to English abstracts (i.e., to 22.3k abstracts, from 4.8k different artists). After applying the Named Entity Recognition, we obtain annotations for each song abstract as depicted in Figure 16. Since our goal is identifying artists in the abstracts, we focus our search on named entities with the types that we found persons or bands to be tagged with: PERSON or ORG (organization).

For our purposes, we therefore consider only those named entities t with tags ORG or PERSON, where t is an artist in the WASABI Song Corpus. We discard all other named entities. After collecting all artist mentions in the song abstracts, we create a graph of artist mentions. Starting with the empty graph G=(V, E), for each artist mention t appearing in one of the song abstracts of an artist A we add both A and t to V. Furthermore, we add edges (A, t) to E. As we consider an undirected graph G, a connection between two artists A and B is established when both artists mention the same other artist t, i.e. if both (A, t) and (B, t) are edges in E. We finally identify corresponding artists A and their mentions t with A in the graph. For example, Cher as artist, Cher as PERSON, and Cher as ORG are all collapsed to one node, *Cher*, in the graph. Since the artist Bob Dylan has an edge to the artist mention Cher as ORG (see Figure 16), we establish a connection in G between the artist Cher and the artist Bob Dylan.

[25] The software can be downloaded at https://spacy.io/. We used the large model *en_core_web_trf* which is based on a transformers architecture.

Figure 17 depicts the most connected artists in the dataset as resulting from the application of our method. The blue circles quantify the total number of other artists mentioning the target artist, for example Bob Dylan has 129 other artists mentioning him in some song abstract.[26] Some artists have more songs with abstracts, which increases their chance to mention other artists, hence appearing as more connected. To account for that bias, we also report a normalized artist connectivity, by discounting the number of song abstracts of an artist. This is illustrated with the interior pink circles inside the blue circles. For example, we find that Bob Dylan is mentioned most often (129 times) but also has a high number of song abstracts (133). Differently, Frank Sinatra is mentioned less often (55 times), but has also far less song abstracts (12). The normalized connectivity of Frank Sinatra is higher than that of Bob Dylan, as visualized by the pink interior circles.

Overall, we can see that the most connected artists are also among the most influential artists in their respective genres and decades: Frank Sinatra (50s/60s), John Lennon (60s/70s), Michael Jackson (80s/90s) as well as newer ones such as Beyoncé and Jay-Z (married to each other) for the 2000s. We find that besides The Rolling Stones there are only artists, not bands, in the most connected list. We speculate that it is more likely for a band member to appear as artist mention, since it is more likely for a band member to co-write, co-compose, co-produce than for a whole band.

At the current stage, we can detect the artist mentions in song abstracts, while we cannot reliably detect yet the relations among the different artists nor the context in which they are mentioned. This is left for future work, and will require linking the artists mentions to knowledge bases, as explained in Section 9.

8.3 Application Scenario C: Linking Lyrics and Real-World Knowledge

The third application scenario, endorsed by archivists from Radio-France, considers the possibility of connecting songs with real world events. For example, during the protest movement against the "yellow jackets" in France, animators of music radio programs have repeatedly requested research for songs about protests, rebellion, anti-government movements, revolution. To support such search, an automated system that given an event finds related songs (i.e., lyrics mentioning or describing such event, or written during a certain historical period) could be developed. As a result of the example query on songs about protests, the U2 song *Sunday Bloody Sunday* could be output (among others), which describes an actual historical event with the same name. While for the general case, the linking between song lyrics and real-world events is complicated (and often there are no specific mentions of them in the lyrics, but some metaphorical reminder), in this specific example, a named entity tagger could

[26] Note that in this figure we only show the artists with the most connections. Most connections from Bob Dylan are not visible as they are connected to not visualized nodes.

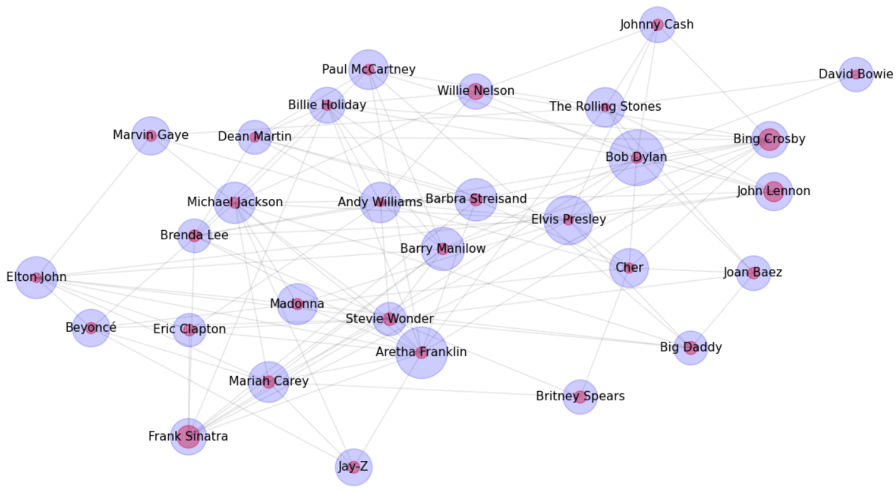


Fig. 17: Connections between artists in the WASABI Song Corpus. Blue circle sizes indicate number of connections, pink interior circles are normalized for song abstracts of artist.

detect the event Bloody Sunday. An entity linker such as Babelfy[27] consequently could link it to the corresponding event (i.e., the 1972 Bloody Sunday incident in Ireland where British troops shot and killed unarmed civil rights protesters: `https://en.wikipedia.org/wiki/Bloody_Sunday_(1972)`). The presence of such entities describing a specific event, or of people renewed to have taken part in specific actions, could be additionally crossed with the results of the topic modeling and emotions recognition methods previously described (Sections 7 and 6). In the direction of enabling such linked applications, we are currently working on representing the WASABI Song Corpus as an RDF Knowledge Graph, as described in the following section.

9 The WASABI RDF Knowledge Graph

The WASABI RDF Knowledge Graph relies upon the Semantic Web standards: RDF provides the conceptual data model and syntaxes, RDFS and OWL provide support for knowledge formalization and reasoning, while SPARQL supports querying RDF graphs.

All the tools, scripts, configuration and mapping files required to generate the knowledge graph, as well as the WASABI ontology, are available from the project's Github repository[28].

[27] `http://babelfy.org`
[28] `https://github.com/micbuffa/WasabiDataset/`

9.1 RDF Knowledge Graph Model, Vocabularies and Formalization

The first version of the WASABI RDF Knowledge Graph, published in late 2020, contains the metadata of the songs, artists and albums described in Section 2 (date and place of recording, producer, genre, record label, composers, instruments, etc.), as well as the lyrics summary (Section 4) and explicitness labels (Section 5). The RDF model relies primarily on two vocabularies: the Music Ontology [51], a rich vocabulary to describe musical metadata, and the WASABI ontology [11] that we designed as an extension of the Music Ontology with respect to specific entities and attributes pertaining to the analysis of song lyrics, such as the detected language and the existence of explicit lyrics in a song. Additionally, the WASABI Knowledge Graph reuses classes and properties from several common metadata vocabularies: Dublin Core Metadata, FOAF, SCOT[29], Schema.org and the DBpedia ontology. Specialized terms were also imported from the Audio Features Ontology[30] and the OMRAS2 Chord Ontology[31].

We recently published version 2.0 of the dataset that relies on an extension of the WASABI ontology to account for the identified emotions and topics of a song (Sections 6 and 7 respectively). A subset of the RDF representation of a song is depicted in Figure 18, with a specific focus on lyrics-related properties. To represent the emotions, we updated the WASABI ontology to model the valence and arousal dimensions of Russel's model as illustrated in Figure 7. Initially, we considered using the EmOCA ontology [6], however its use requires to represent emotions with emoca:hasMinimum and emoca:hasMaximum properties that are not relevant in WASABI where the valence and arousal exist as single values.

To represent social and emotion tags we reused the SCOT[32] (Social Semantic Cloud of Tags) ontology. Tags are ordered collections of terms of type *scot:Tag*, that we represent as RDF lists linked to the song through properties *wsb:social_tags* and *wsb:emotion_tags*. Similarly, a song lyrics summary is an ordered collection of lines, linked to the song with property *wsb:song_summary*. The same principle applies for topics that are ordered, but in addition each topic consists of a bag of words. Hence, the song is associated with an RDF list of topics with property *wsb:topics*, while each topic has multiple words (*wsb:topic_term*). This representation of topics as bags of words differs from common topics modeling where a topic is a single subject, such as in the Ontopic[33] ontology.

[29] SCOT (Social Semantic Cloud of Tags) Ontology: `http://rdfs.org/scot/spec/`
[30] Audio Features Ontology: `http://purl.org/ontology/af/`
[31] OMRAS2 Chord Ontology: `http://purl.org/ontology/chord/`
[32] `http://rdfs.org/scot/spec/`
[33] `http://www.ontologydesignpatterns.org/ont/dul/ontopic.owl`

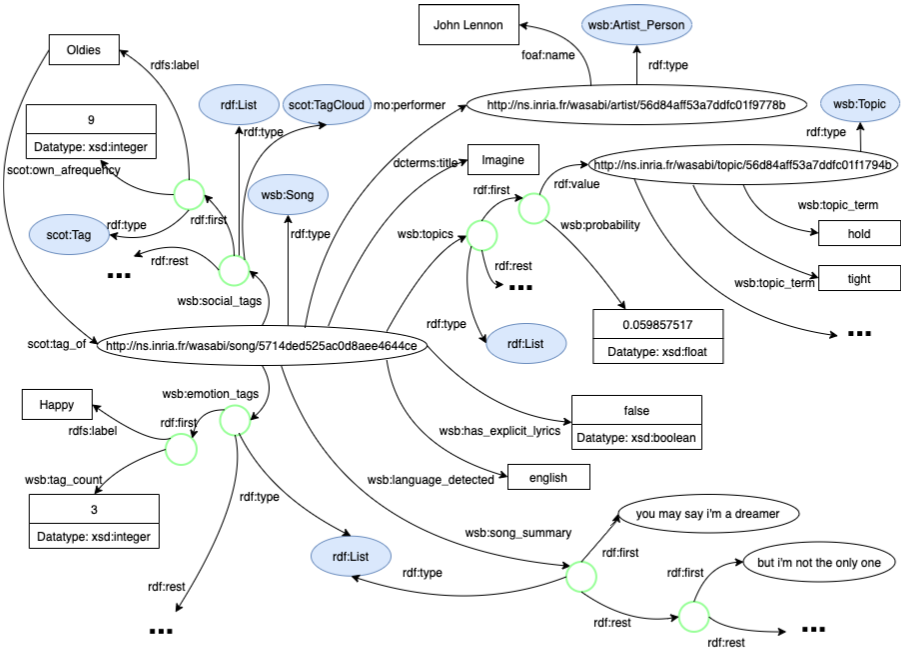


Fig. 18: RDF representation of a song, its artist and some properties related to lyrics analysis. The following namespaces are used: *mo:* Music Ontology, *wsb:* WASABI Ontology, *scot:* SCOT Ontology.).

9.2 Interlinking, Querying and Reasoning on the RDF Knowledge Graph

The 55 million triples RDF dataset is available as a DOI-identified dump that can be downloaded from Zenodo[34], and a public SPARQL endpoint[35] set up to enable the community to query the knowledge graph. Complying with Linked Data good practices, all resources in the knowledge graph, may they denote vocabulary terms, concepts or physical objects, are given URIs that *i)* can be dereferenced to documents describing the resources (selecting RDF or HTML based on content negotiation), and *ii)* can be shared across datasets, making it possible to interlink resources with rich semantic relationships.

In the WASABI RDF Knowledge Graph, all artists, songs and albums are linked to their corresponding web pages in multiple websites. More interestingly, songs and albums are linked to genres named after their DBpedia URIs. Therefore, beyond looking within the WASABI dataset for songs of a given genre, SPARQL federated queries make it possible to query WASABI and DBpedia simultaneously, so as to aggregate songs pertaining to a given set of genres from both sources.

[34] Version 1: `https://doi.org/10.5281/zenodo.4312641`, version 2: `https://doi.org/10.5281/zenodo.5603369`

[35] `http://wasabi.inria.fr/sparql`

We are currently working on broadening these possibilities of simultaneously querying multiple datasets, relying not only on linked resources but also on external knowledge to improve different reasoning tasks. Following the same direction as the work shown in Application Scenario B (Section 8.2), we are currently investigating how DBpedia abstracts (text in natural language providing rich information about artists biography, influences, historical and political contexts, interpretations of a song's meaning and so on) can be used to extract Named Entities and the relationships among them. As shown in Application Scenario C (Section 8.3), being able to link such Named Entities to structured knowledge bases (such as DBpedia and Wikidata) would allow to push forward with the lyrics understanding through advanced reasoning capabilities. To answer the query about the songs pertaining to protests and rebellions, we could retrieve all songs mentioning Wikidata entities that are instances or sub-classes of the *massacre* class, or DBpedia entities that are instances or sub-classes of the *MilitaryConflict* class. Results about all sorts of conflicts would be returned, including e.g. the song *Sunday Bloody Sunday.* Such query can then be specialized to filter undesired entities (and songs).

Furthermore, reasoning can be conveyed by the application of rules implementing specific domain knowledge, meant to infer new facts. For instance, "the influence of one artist on another" can be expressed in a SPARQL query implementing rules such as: "if artist A was a producer of artist B, then A influenced B", or "if artists A and B worked together on album C, then they influenced each other". Again, disambiguating the entities mentioned in lyrics and in the DBpedia abstracts through their linking to knowledge bases (e.g. following the relation `dbo:producer` in DBpedia) would definitely help in the reasoning task.

10 Related Work

This section describes available databases containing songs and lyrics, and summarizes existing work on Natural Language Processing of song lyrics.

10.1 Databases of Songs and Lyrics

The Million Song Dataset (MSD) project[36] [7] is a collection of audio features and metadata for a million contemporary popular music tracks. Such dataset shares some similarities with WASABI with respect to metadata extracted from Web resources (as artist names, tags, years) and audio features, even if at a smaller scale. Given that it mainly focuses on audio data, a complementary dataset providing lyrics of the Million Song dataset was released, called musiXmatch dataset[37]. It consists in a collection of song lyrics in bag-of-words

[36] http://millionsongdataset.com

[37] http://millionsongdataset.com/musixmatch/

(plus stemmed words), associated with MSD tracks. However, no other processing of the lyrics is done, as is the case in our work.

MusicWeb and its successor MusicLynx [2] link music artists within a Web-based application for discovering connections between them and provides a browsing experience using extra-musical relations. The project shares some ideas with WASABI, but works on the artist level, and does not perform analyses on the audio and lyrics content itself. It reuses, for example, MIR metadata from AcousticBrainz.

The WASABI project was built on a broader scope than these projects and mixes a wider set of metadata, including ones from audio and natural language processing of lyrics. In addition, as presented in this paper, it comes with a large set of Web Audio enhanced applications (multitrack player, online virtual instruments and effect, on-demand audio processing, audio player based on extracted, synchronized chords, etc.)

Some of the goals of the DOREMUS project [37] overlap with WASABI, yet in a rather different context. DOREMUS integrates musical metadata from the Bibliothèque nationale de France, Radio France and Philharmonie de Paris. It focuses specifically on classical and traditional music, therefore it is likely that there is little overlap with WASABI that focuses on popular songs. Furthermore, DOREMUS does not deal with audio signal analysis nor lyrics processing as in WASABI, but integrates MIDI resources to achieve music recommendation and automatic playlists generation.

The Listening Experience Database (LED) collects people's music listening experiences as they are reported in documents like diaries, memoirs, letters or oral history recording [1]. It mostly relates to legacy music that has little overlap with WASABI.

The MELD framework [46] supports the publication of musicology articles with multi-modal user interfaces that connect different forms of digital resources such as text, audio, video, images and musical scores. Some development could be undertaken to allow musicologists publish articles that would leverage musical data from the WASABI RDF Knowledge Graph.

The MIDI Linked Data project [40] publishes a large knowledge graph representing in RDF over 300,000 MIDI files. Each one is linked to its DBpedia counterpart, and the RDF model relies on the MIDI ontology and Music Ontology. As such, MIDI Linked Data can be used as a complement of WASABI to allow working on MIDI files together with the audio and text analyses provided by WASABI. It is worth noticing that some MIDI content was used during the evaluation of the chords extraction in WASABI[48, 50, 49] yet not from the MIDI Linked Data project.

The works mentioned above are all about public databases. Let us however notice that companies such as Spotify, GraceNote, Pandora, or Apple Music have sophisticated private knowledge bases of songs and lyrics to feed their search and recommendation algorithms, but there are not publicly available (and mainly rely on audio features).

10.2 Natural Language Processing of Song Lyrics

Lyrics Segmentation. Only a few papers in the literature have focused on the automated detection of the structure of lyrics. Watanabe *et al.* [58] propose the task to automatically identify segment boundaries in lyrics and train a logistic regression model for the task with the repeated pattern and textual features. Mahedero *et al.* [39] report experiments on the use of standard NLP tools for the analysis of music lyrics. Among the tasks they address, for structure extraction they focus on a small sample of lyrics having a clearly recognizable structure (which is not always the case) divided into segments. More recently, Baratè *et al.* [4] describe a semantics-driven approach to the automatic segmentation of song lyrics, and mainly focus on pop/rock music. Their goal is not to label a set of lines in a given way (e.g. verse, chorus), but rather identifying recurrent as well as non-recurrent groups of lines. They propose a rule-based method to estimate such structure labels of segmented lyrics.

Explicit Content Detection. Bergelid [5] consider a dataset of English lyrics to which they apply classical machine learning algorithms. The explicit labels are obtained from Soundtrack Your Brand [38]. They also experiment with adding lyrics metadata to the feature set, such as the artist name, the release year, the music energy level, and the valence/positiveness of a song. Chin *et al.* [19] apply explicit lyrics detection to Korean song texts. They also use tf-idf weighted BOW as lyrics representation and aggregate multiple decision trees via boosting and bagging to classify the lyrics for explicit content. More recently, Kim and Mun [35] proposed a neural network method to create explicit words dictionaries automatically by weighting a vocabulary according to all words' frequencies in the explicit class vs. the clean class, accordingly. They work with a corpus of Korean lyrics.

Emotion Recognition Recently, Delbouys *et al.* [20] address the task of multimodal music mood prediction based on the audio signal and the lyrics of a track. They propose a new model based on deep learning outperforming traditional feature engineering based approaches. Performances are evaluated on their published dataset with associated valence and arousal values which we introduced in Section 6

Xia *et al.* [59] model song texts in a low-dimensional vector space as bags of concepts, the "emotional units"; those are combinations of emotions, modifiers and negations. Yang and Lee [60] leverage the music's emotion annotations from AllMusic which they map to a lower dimensional psychological model of emotion. They train a lyrics emotion classifier and show by qualitative interpretation of an ablated model (decision tree) that the deciding features leading to the classes are intuitively plausible. Hu *et al.* [34] aim to detect emotions in song texts based on Russell's model of mood; rendering emotions continuously in the two dimensions of arousal and valence (positive/negative). They

[38] `https://www.soundtrackyourbrand.com`

analyze each sentence as bag of “emotional units”; they reweight sentences’ emotions by both adverbial modifiers and tense and even consider progressing and adversarial valence in consecutive sentences. Additionally, singing speed is taken into account. With the fully weighted sentences, they perform clustering in the 2D plane of valence and arousal. Although the method is unsupervised at runtime, there are many parameters tuned manually by the authors in this work.

Mihalcea and Strapparava [43] render emotion detection as a multi-label classification problem, songs express intensities of six different basic emotions: anger, disgust, fear, joy, sadness, surprise. Their corpus (100 song texts) has time-aligned lyrics with information on musical key and note progression. Using Mechanical Turk each line of song text is annotated with the six emotions. For emotion classification, they use bags of words and concepts, as musical features key and notes. Their classification results using both modalities, textual and audio features, are significantly improved compared to a single modality.

Topic Modelling Among the works addressing this task for song lyrics, Mahedero *et al.* [39] define five ad hoc topics (Love, Violent, Antiwar, Christian, Drugs) into which they classify their corpus of 500 song texts using supervision. Related, Fell [22] also uses supervision to find bags of genre-specific n-grams. Employing the view from the literature that BOWs define topics, the genre-specific terms can be seen as mixtures of genre-specific topics.

Logan *et al.* [38] apply the unsupervised topic model Probabilistic LSA to their ca. 40k song texts. They learn latent topics for both the lyrics corpus as well as a NYT newspaper corpus (for control) and show that the domain-specific topics slightly improve the performance in their MIR task. While their MIR task performs highly better when using acoustic features, they discover that both methods err differently. Kleedorfer *et al.* [36] apply Non-negative Matrix Factorization (NMF) to ca. 60k song texts and cluster them into 60 topics. They show the so discovered topics to be intrinsically meaningful.

Sterckx [53] have worked on topic modeling of a large-scale lyrics corpus of 1M songs. They build models using Latent Dirichlet allocation with topic counts between 60 and 240 and show that the 60 topics model gives a good trade-off between topic coverage and topic redundancy. Since popular topic models such as LDA represent topics as weighted bags of words, these topics are not immediately interpretable. This gives rise to the need of an automatic labeling of topics with smaller labels. A recent approach [8] relates the topical BOWs with titles of Wikipedia articles in a two step procedure: first, candidates are generated, then ranked.

For the topic of network analysis in the song lyrics domain, Atherton *et al.* [3] have analyzed how different song writers have influenced one another by investigating how trigrams in the lyrics are initially used by one writer and later in time reproduced by other writers.

11 Conclusion

In this paper we have described the WASABI Song Corpus, focusing in particular on the lyrics annotations resulting from the applications of the methods we proposed to extract relevant information from the lyrics. So far, lyrics annotations concern their structure segmentation, their topic, the explicitness of the lyrics content, the summary of a song and the emotions conveyed. We motivated using our corpus by presenting three different application scenarios. While parts of them are not yet fully implemented, we described our progress in transforming our dataset into a knowledge graph, the WASABI RDF Knowledge Graph, which will be vital both in these implementations and in enabling further interesting application scenarios in future work.

Acknowledgements This work is partly funded by the French Research National Agency (ANR) under the WASABI project (contract ANR-16-CE23-0017-01).

References

1. Adamou, A., Brown, S., Barlow, H., Allocca, C., d'Aquin, M.: Crowdsourcing linked data on listening experiences through reuse and enhancement of library data. International Journal on Digital Libraries **20**(1), 61–79 (2019)
2. Allik, A., Thalmann, F., Sandler, M.: MusicLynx: Exploring music through artist similarity graphs. In: Companion Proc. (Dev. Track) The Web Conf. (WWW 2018) (2018)
3. Atherton, J., Kaneshiro, B.: I said it first: Topological analysis of lyrical influence networks. In: ISMIR, pp. 654–660 (2016)
4. Baratè, A., Ludovico, L.A., Santucci, E.: A semantics-driven approach to lyrics segmentation. In: 2013 8th International Workshop on Semantic and Social Media Adaptation and Personalization, pp. 73–79 (2013). DOI 10.1109/SMAP.2013.15
5. Bergelid, L.: Classification of explicit music content using lyrics and music metadata (2018)
6. Berthelon, F., Sander, P.: Emotion Ontology for Context Awareness. In: Coginfocom 2013 - 4th IEEE Conference on Cognitive Infocommunicaitons. Budapest, Hungary (2013). URL `https://hal.archives-ouvertes.fr/hal-00908543`
7. Bertin-Mahieux, T., Ellis, D.P., Whitman, B., Lamere, P.: The million song dataset. In: Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011) (2011)
8. Bhatia, S., Lau, J.H., Baldwin, T.: Automatic labelling of topics with neural embeddings. arXiv preprint arXiv:1612.05340 (2016)
9. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of machine Learning research **3**(Jan), 993–1022 (2003)
10. Brackett, D.: Interpreting Popular Music. Cambridge University Press (1995). URL `https://books.google.fr/books?id=yHniAAAAMAAJ`
11. Buffa, M., Cabrio, E., Fell, M., Gandon, F., Giboin, A., Hennequin, R., Michel, F., Pauwels, J., Pellerin, G., Tikat, M., Winckler, M.: The WASABI dataset: cultural, lyrics and audio analysis metadata about 2 million popular commercially released songs. In: Proceedings of ESWC 2021 (to be published) (2021)
12. Buffa, M., Lebrun, J.: Real time tube guitar amplifier simulation using webaudio. In: Proc. 3rd Web Audio Conference (WAC 2017) (2017)
13. Buffa, M., Lebrun, J.: Web audio guitar tube amplifier vs native simulations. In: Proc. 3rd Web Audio Conf. (WAC 2017) (2017)
14. Buffa, M., Lebrun, J., Kleimola, J., Letz, S., et al.: Towards an open web audio plugin standard. In: Companion Proceedings of the The Web Conference 2018, pp. 759–766. International World Wide Web Conferences Steering Committee (2018)

15. Buffa, M., Lebrun, J., Pauwels, J., Pellerin, G.: A 2 Million Commercial Song Interactive Navigator. In: WAC 2019 - 5th WebAudio Conference 2019. Trondheim, Norway (2019). URL https://hal.inria.fr/hal-02366730
16. Buffa, M., Lebrun, J., Pellerin, G., Letz, S.: Webaudio plugins in daws and for live performance. In: 14th International Symposium on Computer Music Multidisciplinary Research (CMMR'19) (2019)
17. Çano, E., Morisio, M.: Music mood dataset creation based on last.fm tags. In: 2017 International Conference on Artificial Intelligence and Applications, Vienna Austria (May 2017). DOI 10.5121/csit.2017.70603
18. Chatterjee, A., Narahari, K.N., Joshi, M., Agrawal, P.: Semeval-2019 task 3: Emocontext contextual emotion detection in text. In: Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 39–48 (2019)
19. Chin, H., Kim, J., Kim, Y., Shin, J., Yi, M.Y.: Explicit content detection in music lyrics using machine learning. In: 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 517–521. IEEE (2018)
20. Delbouys, R., Hennequin, R., Piccoli, F., Royo-Letelier, J., Moussallam, M.: Music mood detection based on audio and lyrics with deep neural net. arXiv preprint arXiv:1809.07276 (2018)
21. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
22. Fell, M.: Lyrics classification. In: Master's thesis, Saarland University, Germany, 2014. (2014)
23. Fell, M.: Natural language processing for music information retrieval : deep analysis of lyrics structure and content. Theses, Université Côte d'Azur (2020). URL https://tel.archives-ouvertes.fr/tel-02587910
24. Fell, M., Cabrio, E., Corazza, M., Gandon, F.: Comparing Automated Methods to Detect Explicit Content in Song Lyrics. In: RANLP 2019 - Recent Advances in Natural Language Processing. Varna, Bulgaria (2019). URL https://hal.archives-ouvertes.fr/hal-02281137
25. Fell, M., Cabrio, E., Gandon, F., Giboin, A.: Song lyrics summarization inspired by audio thumbnailing. In: RANLP 2019 - Recent Advances in Natural Language Processing (RANLP). Varna, Bulgaria (2019). URL https://hal.archives-ouvertes.fr/hal-02281138
26. Fell, M., Cabrio, E., Korfed, E., Buffa, M., Gandon, F.: Love me, love me, say (and write!) that you love me: Enriching the WASABI song corpus with lyrics annotations. In: Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020, pp. 2138–2147 (2020). URL https://www.aclweb.org/anthology/2020.lrec-1.262/
27. Fell, M., Nechaev, Y., Cabrio, E., Gandon, F.: Lyrics Segmentation: Textual Macrostructure Detection using Convolutions. In: Conference on Computational Linguistics (COLING), pp. 2044–2054. Santa Fe, New Mexico, United States (2018). URL https://hal.archives-ouvertes.fr/hal-01883561
28. Fell, M., Sporleder, C.: Lyrics-based analysis and classification of music. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 620–631 (2014)
29. Fell, M., Yaroslav, N., Gabriel, M.B., Cabrio, E., Gandon, F., Peeters, G.: Lyrics segmentation via bimodal text-audio representation. Natural Language Engineering (2021, to appear)
30. Fillon, T., Simonnot, J., Mifune, M.F., Khoury, S., Pellerin, G., Le Coz, M.: Telemeta: An open-source web framework for ethnomusicological audio archives management and automatic analysis. In: Proceedings of the 1st International Workshop on Digital Libraries for Musicology, pp. 1–8. ACM (2014)
31. Hennequin, R., Khlif, A., Voituret, F., Moussallam, M.: Spleeter: A fast and state-of-the art music source separation tool with pre-trained models. Late-Breaking/Demo ISMIR 2019 (2019). Deezer Research
32. Honnibal, M., Montani, I., Van Landeghem, S., Boyd, A.: spaCy: Industrial-strength Natural Language Processing in Python (2020). DOI 10.5281/zenodo.1212303. URL https://doi.org/10.5281/zenodo.1212303

33. Hu, X., Downie, J.S., Ehmann, A.F.: Lyric text mining in music mood classification. American music **183**(5,049), 2–209 (2009)
34. Hu, Y., Chen, X., Yang, D.: Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. In: ISMIR (2009)
35. Kim, J., Mun, Y.Y.: A hybrid modeling approach for an automated lyrics-rating system for adolescents. In: European Conference on Information Retrieval, pp. 779–786. Springer (2019)
36. Kleedorfer, F., Knees, P., Pohle, T.: Oh oh oh whoah! towards automatic topic detection in song lyrics. In: ISMIR (2008)
37. Lisena, P., Achichi, M., Choffé, P., Cecconi, C., Todorov, K., Jacquemin, B., Troncy, R.: Improving (re-) usability of musical datasets: An overview of the doremus project. Bibliothek Forschung und Praxis **42**(2), 194–205 (2018)
38. Logan, B., Kositsky, A., Moreno, P.: Semantic analysis of song lyrics. In: 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763), vol. 2, pp. 827–830 Vol.2 (2004). DOI 10.1109/ICME.2004.1394328
39. Mahedero, J.P.G., Martínez, A., Cano, P., Koppenberger, M., Gouyon, F.: Natural language processing of lyrics. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, MULTIMEDIA '05, pp. 475–478. ACM, New York, NY, USA (2005). DOI 10.1145/1101149.1101255. URL `http://doi.acm.org/10.1145/1101149.1101255`
40. Meroño-Peñuela, A., Hoekstra, R., Gangemi, A., Bloem, P., de Valk, R., Stringer, B., Janssen, B., de Boer, V., Allik, A., Schlobach, S., et al.: The midi linked data cloud. In: International Semantic Web Conference, pp. 156–164. Springer (2017)
41. Meseguer-Brocal, G., Peeters, G., Pellerin, G., Buffa, M., Cabrio, E., Faron Zucker, C., Giboin, A., Mirbel, I., Hennequin, R., Moussallam, M., Piccoli, F., Fillon, T.: WASABI: a Two Million Song Database Project with Audio and Cultural Metadata plus WebAudio enhanced Client Applications. In: Web Audio Conference 2017 – Collaborative Audio #WAC2017. Queen Mary University of London, London, United Kingdom (2017)
42. Michel Buffa Maroua Tikat, M.W.: Interactive multimedia visualization for exploring and fixing a multi-dimensional metadata base of popular musics. In: Proceedings of the MEPDaW Workshop, ISWC 2021 (2021)
43. Mihalcea, R., Strapparava, C.: Lyrics, music, and emotions. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 590–599. Association for Computational Linguistics, Jeju Island, Korea (2012). URL `https://www.aclweb.org/anthology/D12-1054`
44. Mohammad, S.: Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 174–184 (2018)
45. Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: Semeval-2018 task 1: Affect in tweets. In: Proceedings of the 12th international workshop on semantic evaluation, pp. 1–17 (2018)
46. Page, K.R., Lewis, D., Weigl, D.M.: Meld: a linked data framework for multimedia access to music digital libraries. In: 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL), pp. 434–435. IEEE (2019)
47. Parisi, L., Francia, S., Olivastri, S., Tavella, M.S.: Exploiting synchronized lyrics and vocal features for music emotion detection. CoRR **abs/1901.04831** (2019). URL `http://arxiv.org/abs/1901.04831`
48. Pauwels, J., O'Hanlon, K., Fazekas, G., Sandler, M.: Confidence measures and their applications in music labelling systems based on hidden Markov models. In: Proc. 18th Int. Soc. Music Information Retrieval (ISMIR 2017) (2017)
49. Pauwels, J., Sandler, M.: A web-based system for suggesting new practice material to music learners based on chord content. In: Joint Proc. 24th ACM IUI Workshops (IUI2019) (2019)
50. Pauwels, J., Xambó, A., Roma, G., Barthet, M., Fazekas, G.: Exploring real-time visualisations to support chord learning with a large music collection. In: Proc. 4th Web Audio Conf. (WAC 2018) (2018)
51. Raimond, Y., Abdallah, S., Sandler, M., Giasson, F.: The Music Ontology. In: Proceedings of the 8th ISMIR Conf., pp. 417–422 (2007)

52. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology **39**(6), 1161 (1980)
53. Sterckx, L.: Topic detection in a million songs. Ph.D. thesis, PhD thesis, Ghent University (2014)
54. Stöter, F.R., Uhlich, S., Liutkus, A., Mitsufuji, Y.: Open-unmix-a reference implementation for music source separation. Journal of Open Source Software (2019)
55. Tagg, P.: Analysing popular music: theory, method and practice. Popular Music **2**, 37–67 (1982). DOI 10.1017/S0261143000001227
56. Vanni, L., Ducoffe, M., Aguilar, C., Precioso, F., Mayaffre, D.: Textual deconvolution saliency (tds): a deep tool box for linguistic analysis. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 548–557 (2018)
57. Warriner, A.B., Kuperman, V., Brysbaert, M.: Norms of valence, arousal, and dominance for 13,915 english lemmas. Behavior research methods **45**(4), 1191–1207 (2013)
58. Watanabe, K., Matsubayashi, Y., Orita, N., Okazaki, N., Inui, K., Fukayama, S., Nakano, T., Smith, J., Goto, M.: Modeling discourse segments in lyrics using repeated patterns. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 1959–1969 (2016)
59. Xia, Y., Wang, L., Wong, K.F., Xu, M.: Sentiment vector space model for lyric-based song sentiment classification. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, HLT-Short '08, pp. 133–136. Association for Computational Linguistics, Stroudsburg, PA, USA (2008). URL `http://dl.acm.org/citation.cfm?id=1557690.1557725`
60. Yang, D., Lee, W.: Music emotion identification from lyrics. In: 2009 11th IEEE International Symposium on Multimedia, pp. 624–629 (2009). DOI 10.1109/ISM.2009.123